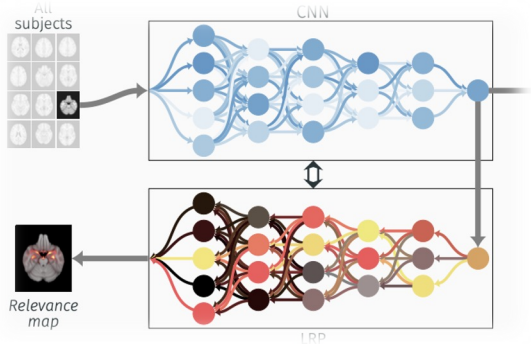
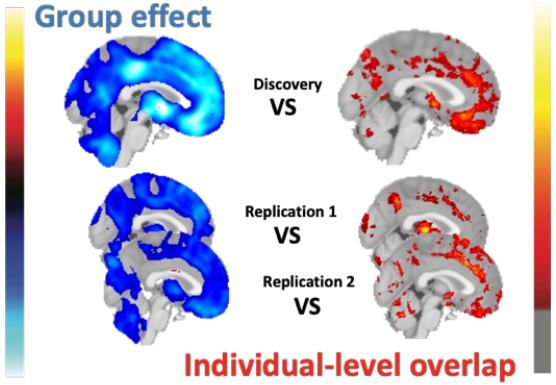
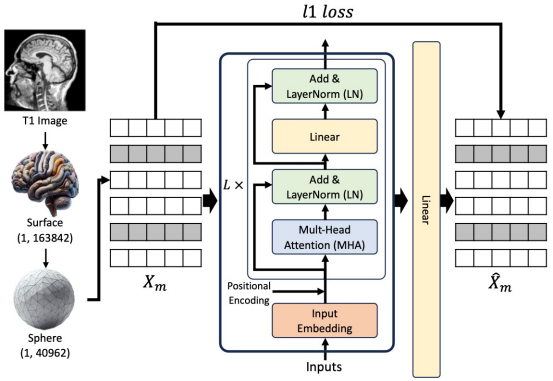




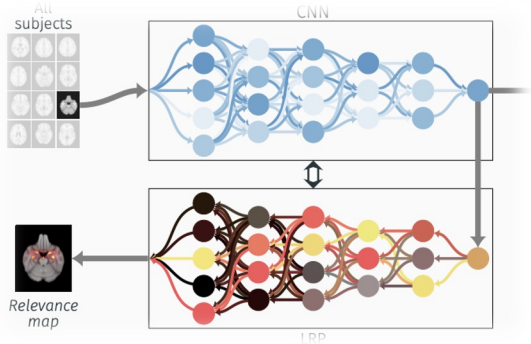


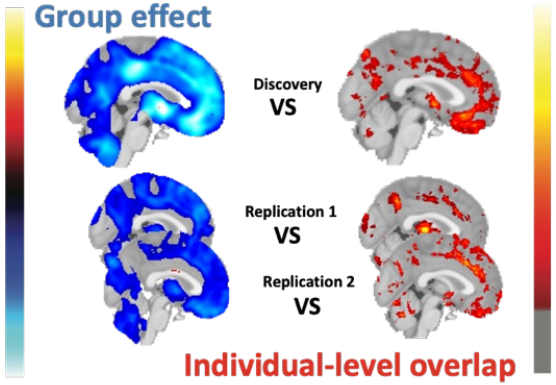
Predicting Brain Age from MRI Scans via Deep Learning

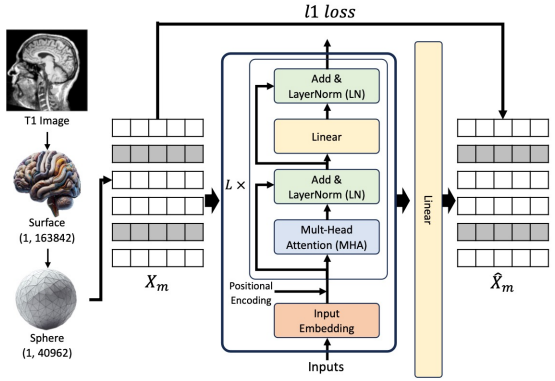
Francesco Sammarco

Ph.d. @ *Laboratory for Mental Health Mapping*

University of Jena//Tübingen







Outline Presentation

Part 1: EDA

Part 2: Regression

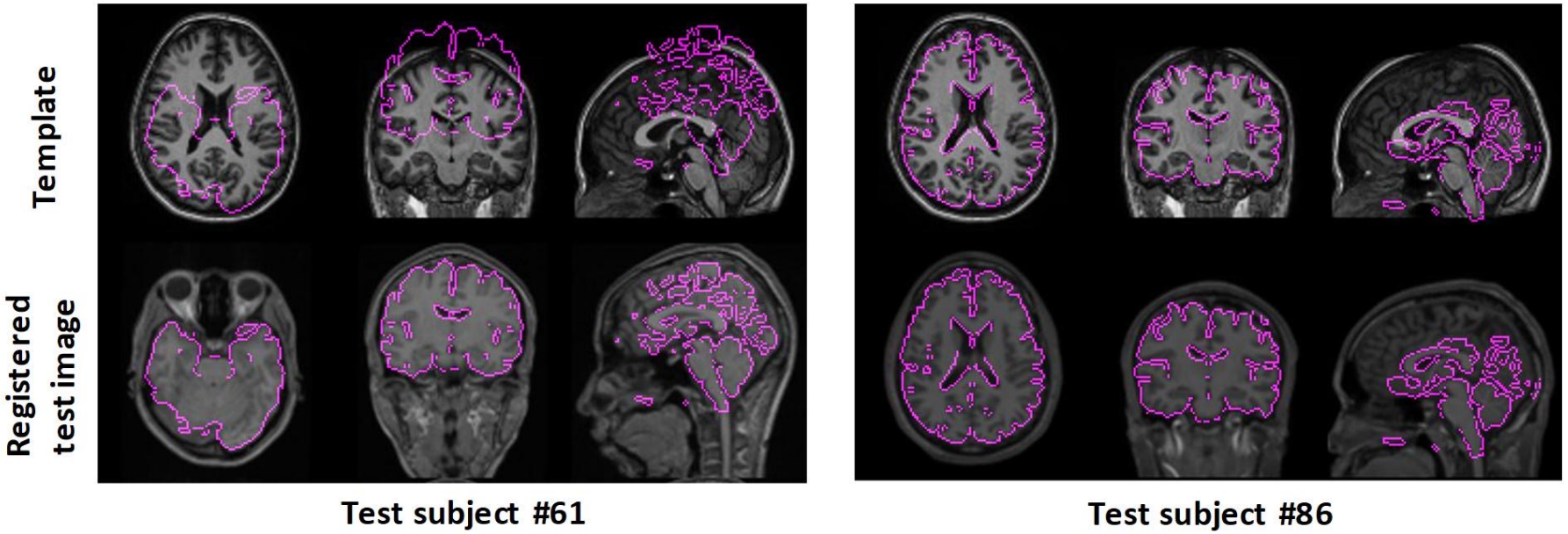
Part 3: Deep Learning

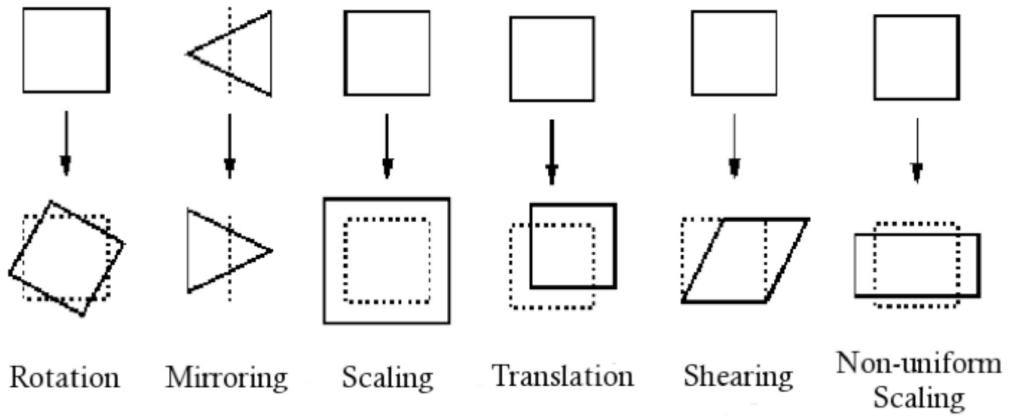
Part 4: Explainability

Part 5: Personal Projects

https://mhm-lab.github.io

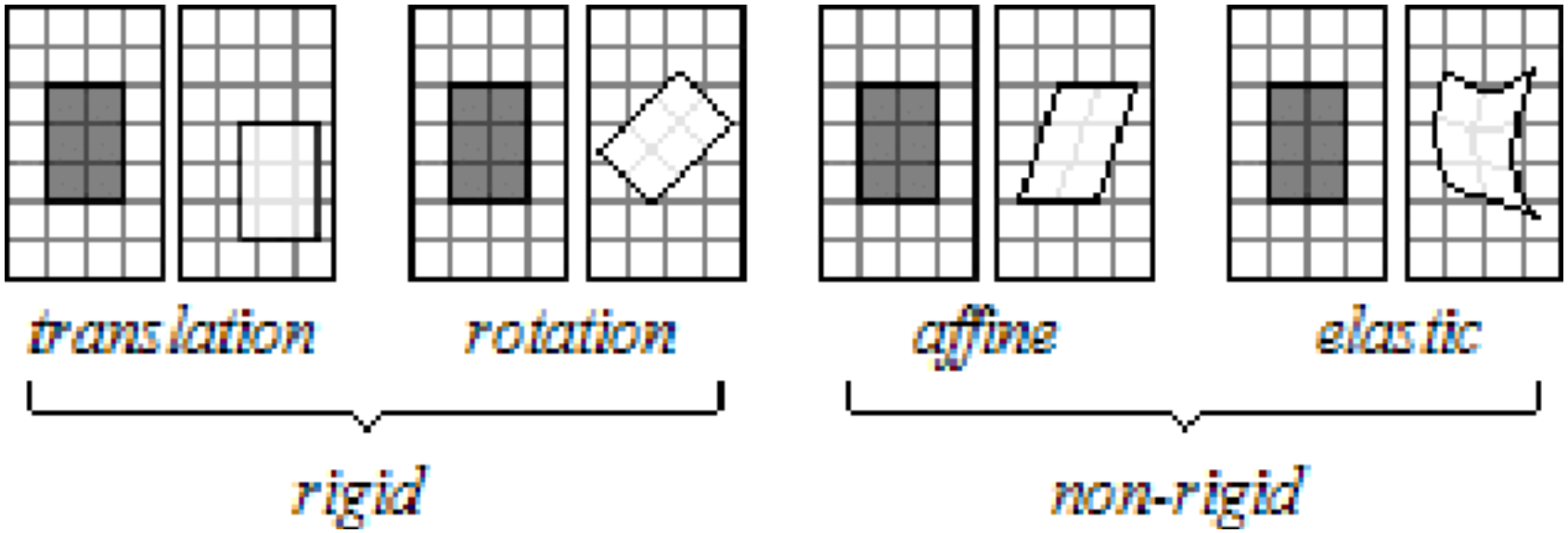
EDA: Registration of MRI images



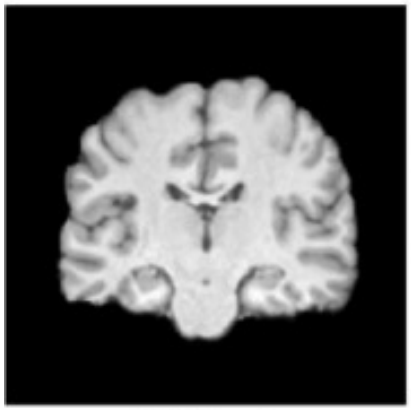


Mental Health Mapping

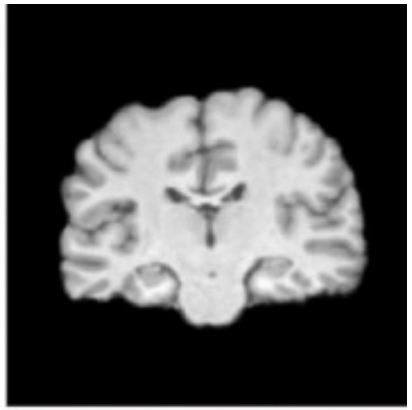
EDA: Registration of MRI images



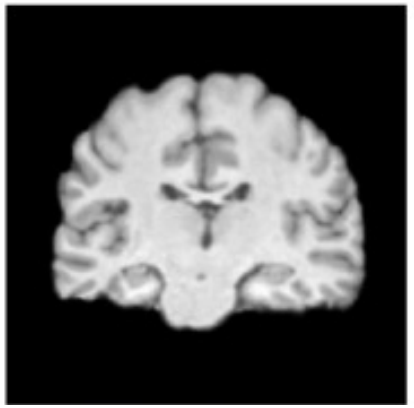
EDA: Registration of MRI images



Original



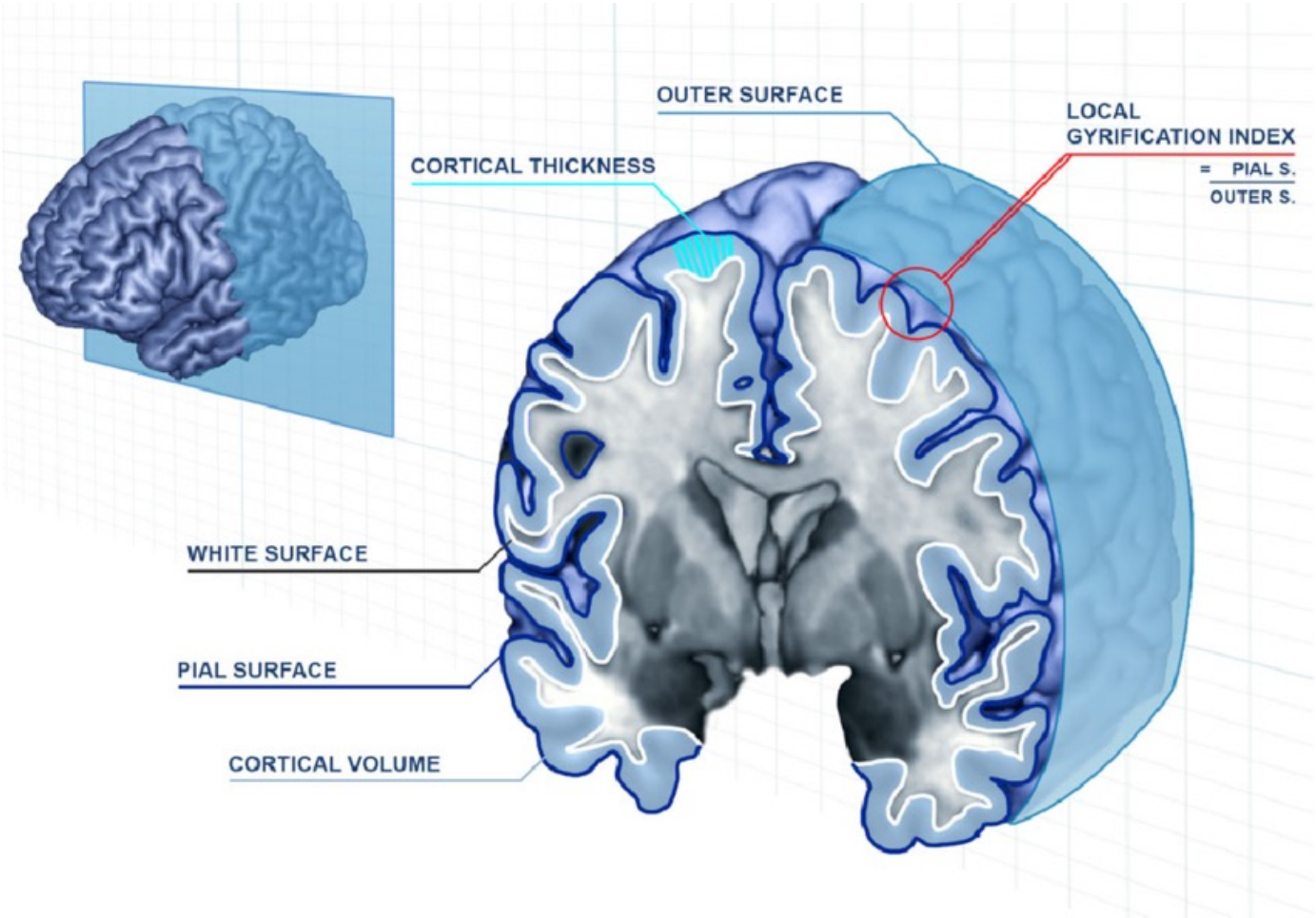
Affine



Elastic

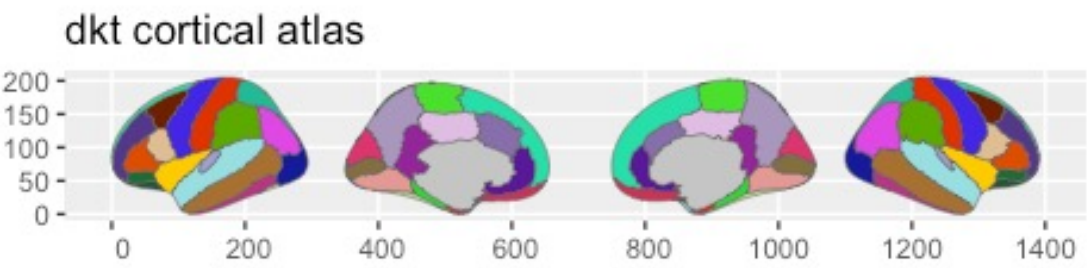
Mental Health Mapping

EDA: Corthical thickness



Mental Health Mapping

EDA: The DKT Atlas



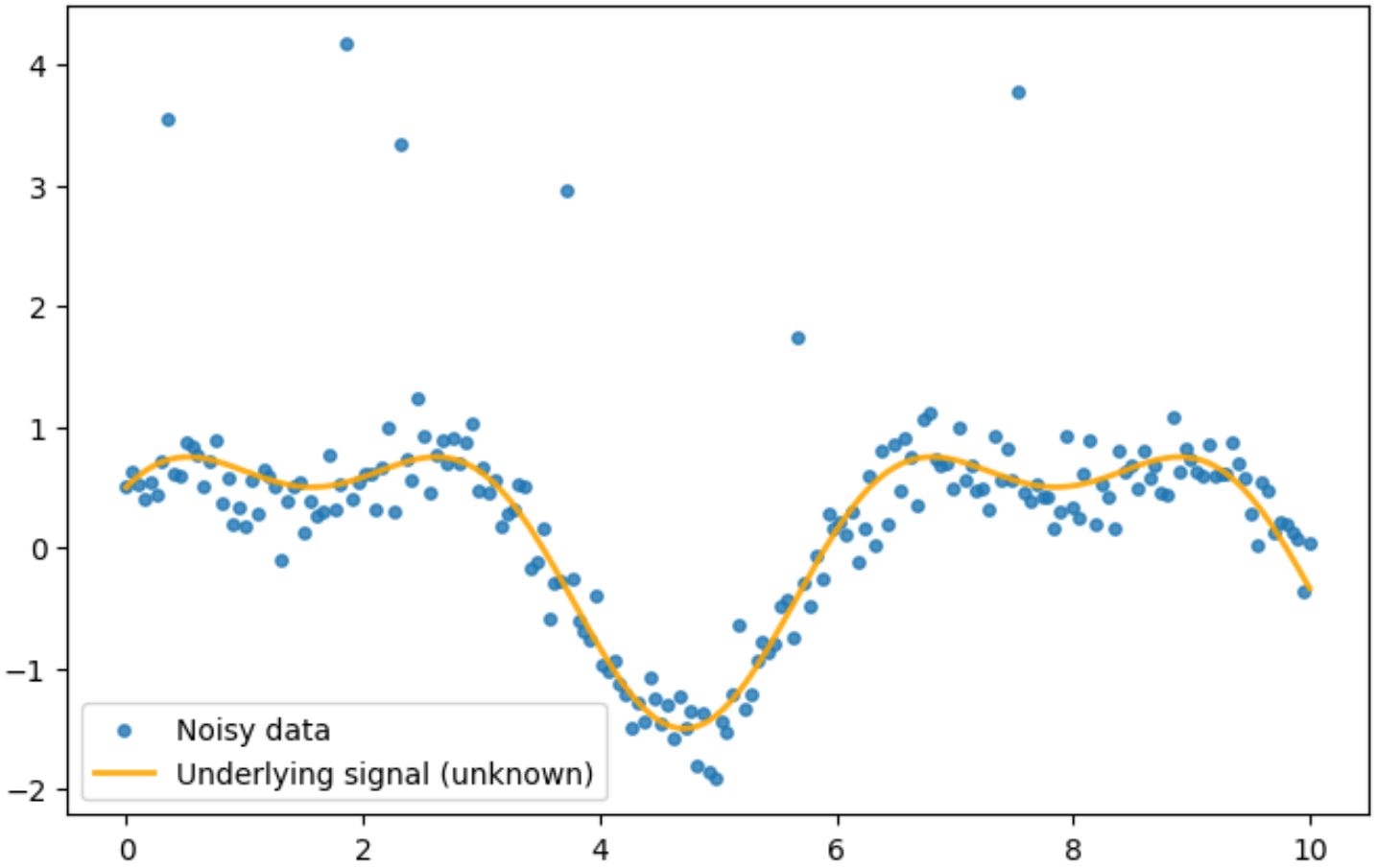
caudal anterior cingulate
caudal middle frontal
cuneus
entorhinal
fusiform
inferior parietal
inferior temporal
insula
isthmus cingulate
lateral occipital
lateral orbitofrontal
lingual
medial orbitofrontal
middle temporal
paracentral
parahippocampal
pars opercularis
pars orbitalis
pars triangularis
pericalcarine
postcentral
posterior cingulate
precentral
precuneus
rostral anterior cingulate
rostral middle frontal
superior frontal
superior parietal
superior temporal
supramarginal
transverse temporal
NA

Mental Health Mapping

Regression: Locally estimated scatterplot smoothing

Let's consider a set of randomly sampled points with the function:

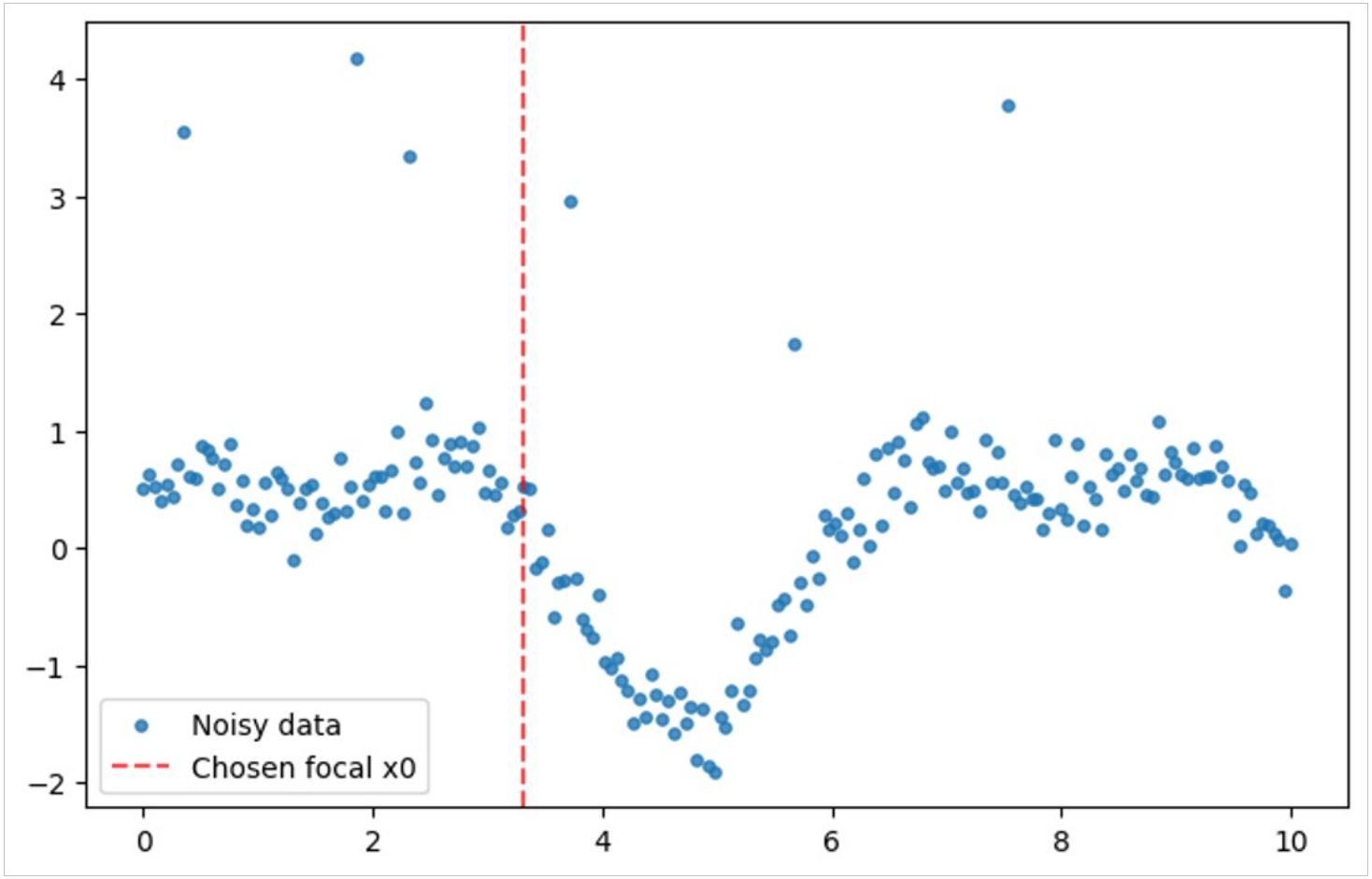
$$\sin x + \tfrac{1}{2}\cos 2x + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \tfrac{1}{4})$$



Mental Health Mapping

Regression: Locally estimated scatterplot smoothing

We consider a focal point x_0:



Regression: Locally estimated scatterplot smoothing

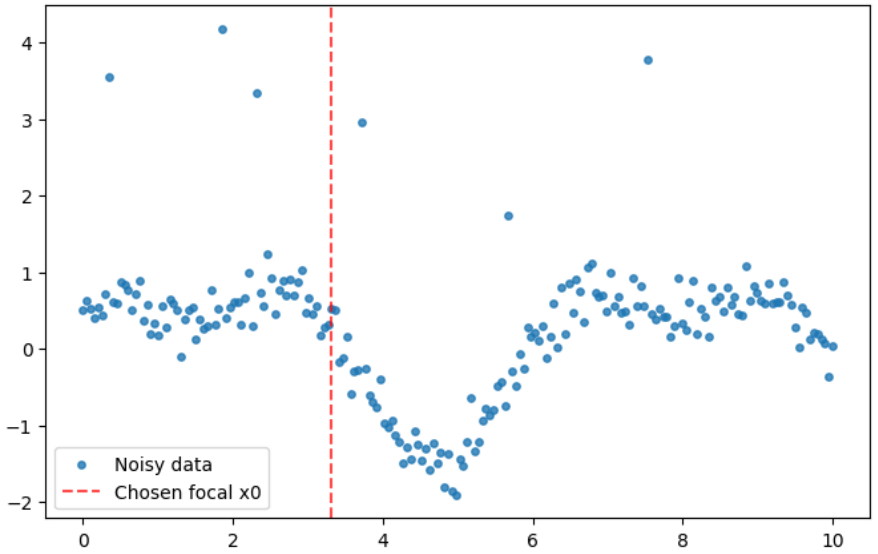
We assign all points around x_0 a weight function W. A typical choice is the tricube function:

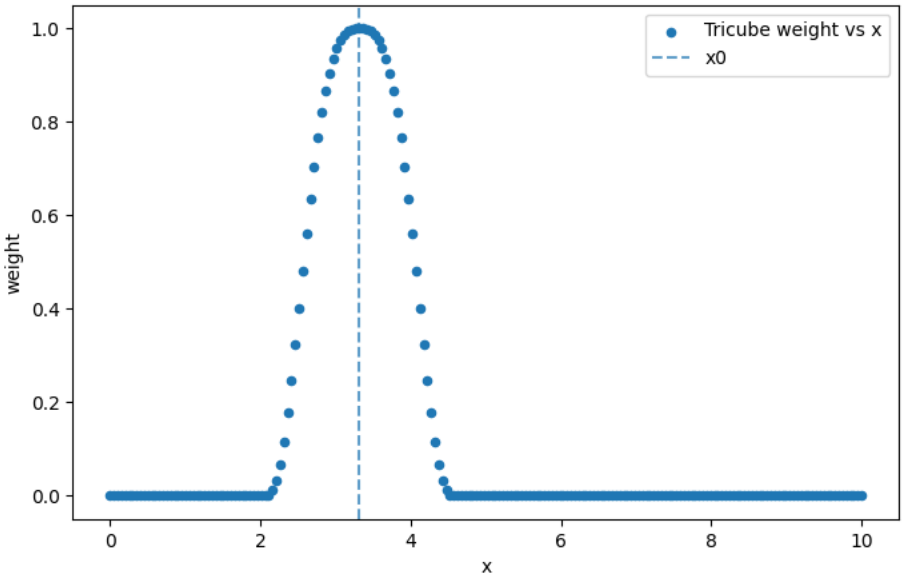
$$W(u) = (1 - |u|^3)^3, \quad |u| < 1$$

Applying it "around x_0" means considering u as:

$$u(x) = \frac{x_0 - x}{h}$$

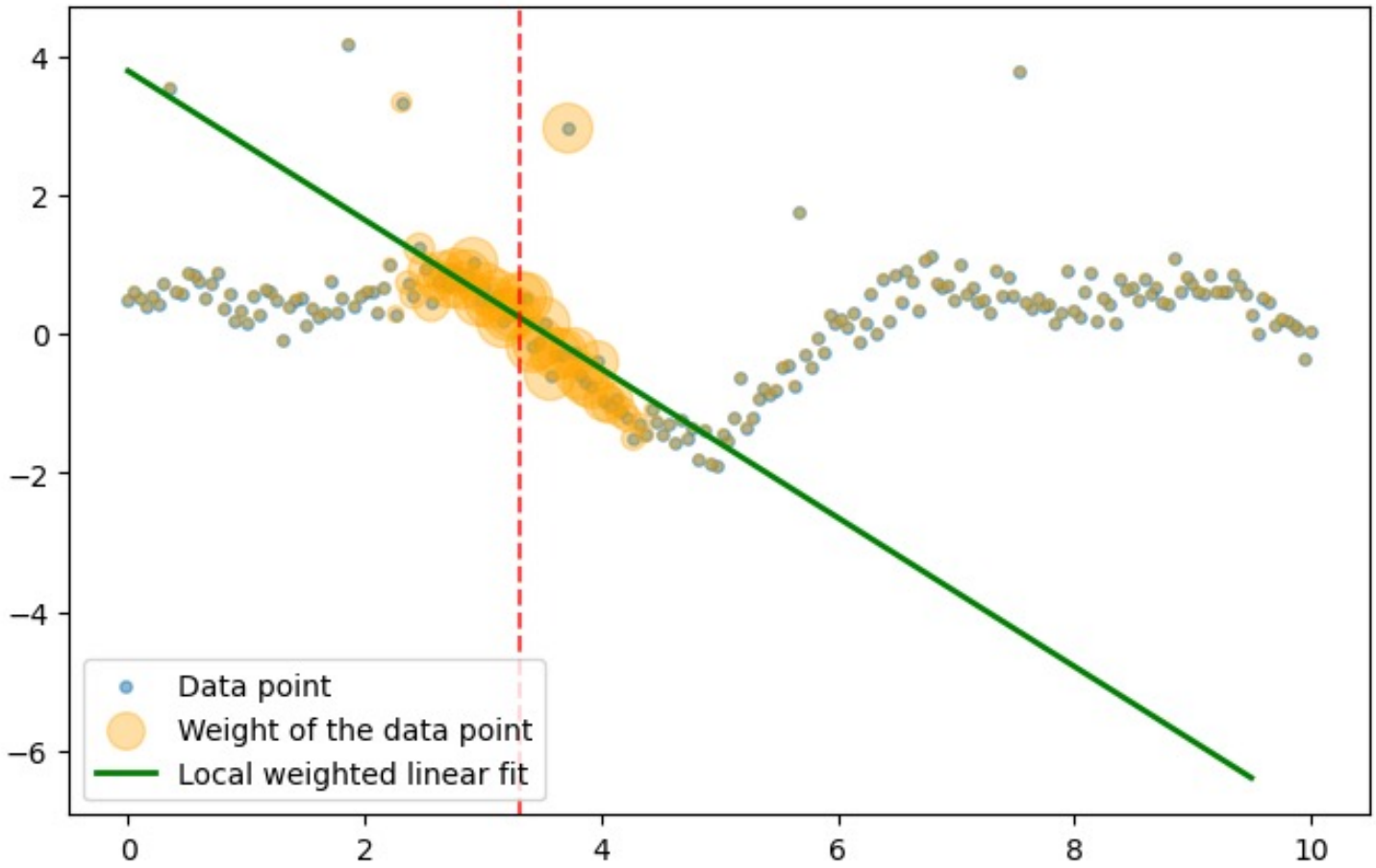
Where h is a bandwidth parameter.





Regression: Locally estimated scatterplot smoothing

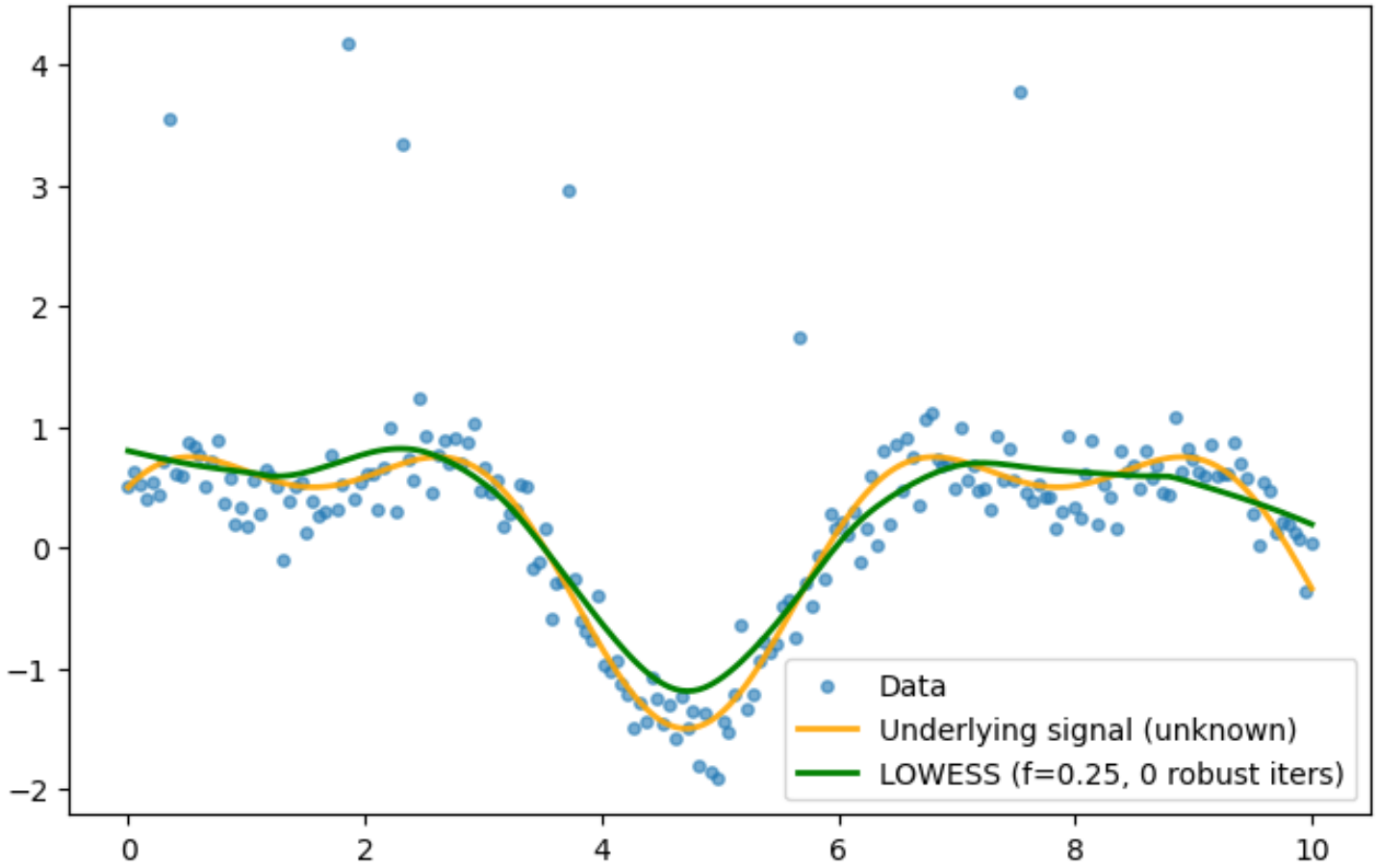
A local weighted regression model is fitted on the points, with weight from the function:



Mental Health Mapping

Regression: Locally estimated scatterplot smoothing

Fit this model for all values of x. The final model is composed by all the intercepts.



Mental Health Mapping

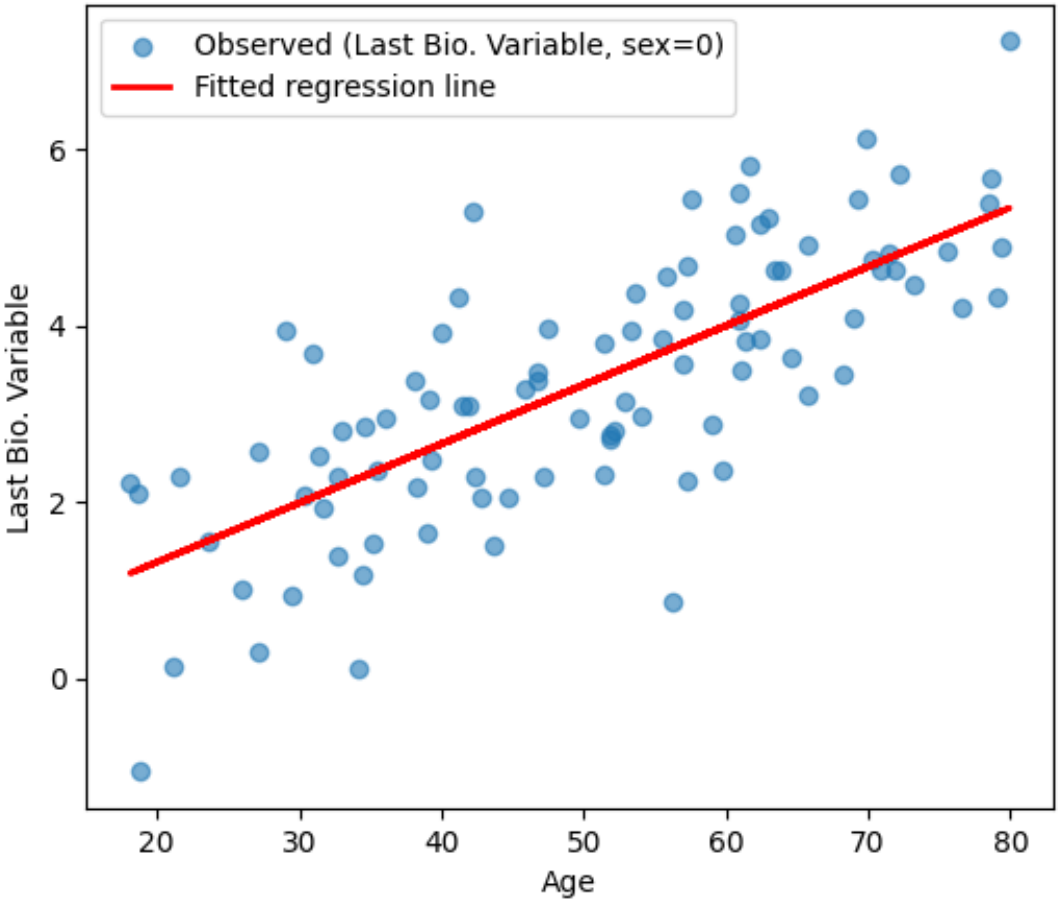
Regression: COMBAT

COMBAT supposes that the generic observation is structured by a biological fit and a "non-biological" fit:

$$y_{ij} = (\text{biological fit})_{ij} \qquad \underbrace{\gamma_{bj} + \epsilon_{ij}}_{\text{batch effect + noise}}$$

Regression: COMBAT

The idea is to separate the “biological” contribution from the “external contribution”. First we fit each biological variable via the biological variables.



Mental Health Mapping

Regression: COMBAT

Now, we consider the difference not explained by the variables, i.e. the residuals. When separating each residual per “non-biological” variable, we can estimate mean and variance.

$$\hat{\gamma_{b,j}} = \frac{\sum_{i=0}^{N_b} y_j - \hat{y_j}}{N_b}$$

$$\hat{\sigma_{b,j}} = \frac{\sum_{i=0}^{N_b} |(y_j - \hat{y_j}) - \hat{\gamma_{b,j}}|^2}{N_b - 1}$$

Mental Health Mapping

Regression: COMBAT

COMBAT assumes that each of the parameter obtained as before is drawn from a prior distribution. For the "mean" (called *location effect*):

$$\gamma_{b,j} \sim \mathcal{N}(\gamma_b, \tau_b)$$

The "standard deviation" (called *scale effect*) is supposed to be drawn by the so called *Inverse Gamma* function:

$$\sigma_{b,j} \sim \frac{\beta_b^{\alpha_b}}{\Gamma(\alpha_b)} x^{-(\alpha_b+1)} \exp\left(-\frac{\beta_b}{x}\right), \quad x > 0$$

Mental Health Mapping

Regression: COMBAT

Doing so allows the method to turn the noisy per-feature estimates into *posterior "shrunken" estimates*. Basically, we estimate the posterior with the Bayes rule:

$$p(\gamma_{bj}, \sigma_{bj} \mid \text{data}) \propto p(\text{data} \mid \gamma_{bj}, \sigma_{bj}) \cdot p(\gamma_{bj}) \cdot p(\sigma_{bj})$$

From the joint distribution, we "marginalize" the single parameter distribution. Since the priors are conjugate, we get again distribution of the same type of the priors.

From these marginals, we compute the *expectation* of our parameters given the data.

$$\mathbb{E}[\sigma_{b,j} \mid \text{data}] = \frac{b_b + \frac{1}{2}(n_b - 1)\hat{\sigma}_{b,j}}{a_b + \frac{n_b}{2} - 1} \qquad \mathbb{E}[\gamma_{b,j} \mid \text{data}] = \frac{\frac{n_b}{\sigma_{b,j}}\gamma_{b,j} + \frac{1}{\tau_b^2}\bar{\gamma}_b}{\frac{n_b}{\mathbb{E}[\sigma_{b,j}|\text{data}]} + \frac{1}{\tau_b^2}}$$

Mental Health Mapping

Regression: COMBAT

Doing so allows the method to turn the noisy per-feature estimates into *posterior "shrunken" estimates*. Basically, we estimate the posterior with the Bayes rule:

$$p(\gamma_{bj}, \sigma_{bj} \mid \text{data}) \propto p(\text{data} \mid \gamma_{bj}, \sigma_{bj}) \cdot p(\gamma_{bj}) \cdot p(\sigma_{bj})$$

From the joint distribution, we "marginalize" the single parameter distribution. Since the priors are conjugate, we get again distribution of the same type of the priors.

From these marginals, we compute the *expectation* of our parameters given the data.

$$\mathbb{E}[\sigma_{b,j} \mid \text{data}] = \frac{b_b + \frac{1}{2}(n_b - 1)\hat{\sigma}_{b,j}}{a_b + \frac{n_b}{2} - 1} \qquad \mathbb{E}[\gamma_{b,j} \mid \text{data}] = \frac{\frac{n_b}{\sigma_{b,j}}\gamma_{b,j} + \frac{1}{\tau_b^2}\bar{\gamma}_b}{\frac{n_b}{\mathbb{E}[\sigma_{b,j}|\text{data}]} + \frac{1}{\tau_b^2}}$$

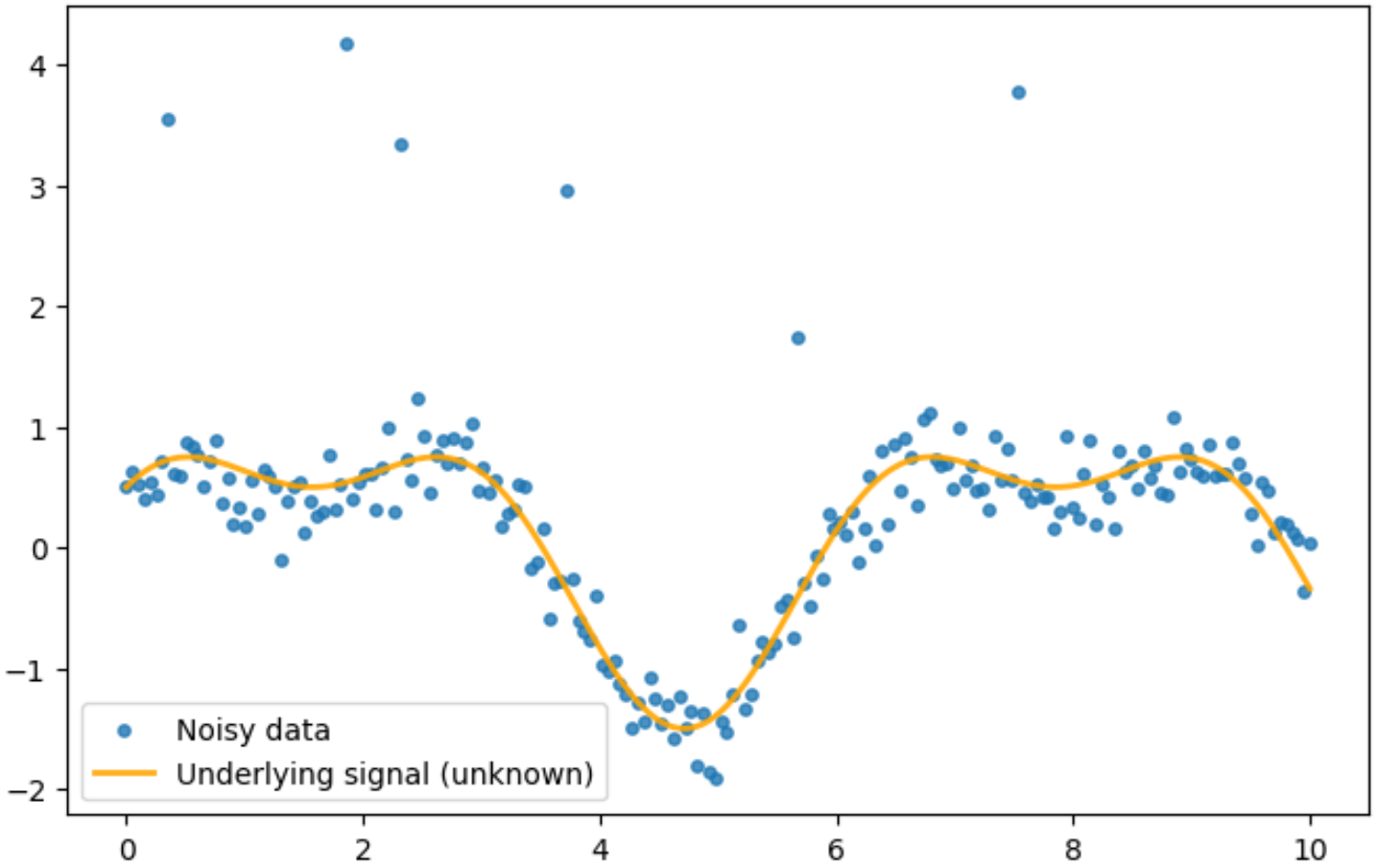
This are used to "shrink" the residuals and obtain the final scaled y values.

Mental Health Mapping

Regression: Kernel Ridge Regression

In Kernel regression, the idea is always to fit a set of points. Suppose again:

$$\sin x + \tfrac{1}{2}\cos 2x + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \tfrac{1}{4})$$



Mental Health Mapping

Regression: Kernel Ridge Regression

Kernel regression allows to model a set of complex features of the inputs by expressing them in terms of similarity of the inputs (the so called *kernel trick*):

$$f(x) = \sum_{i=1}^{n} \alpha_i \langle \phi(x_i), \phi(x) \rangle = \sum_{i=1}^{n} \alpha_i \, k(x_i, x),$$

$$k(x, x') = \phi(x)^\top \phi(x')$$

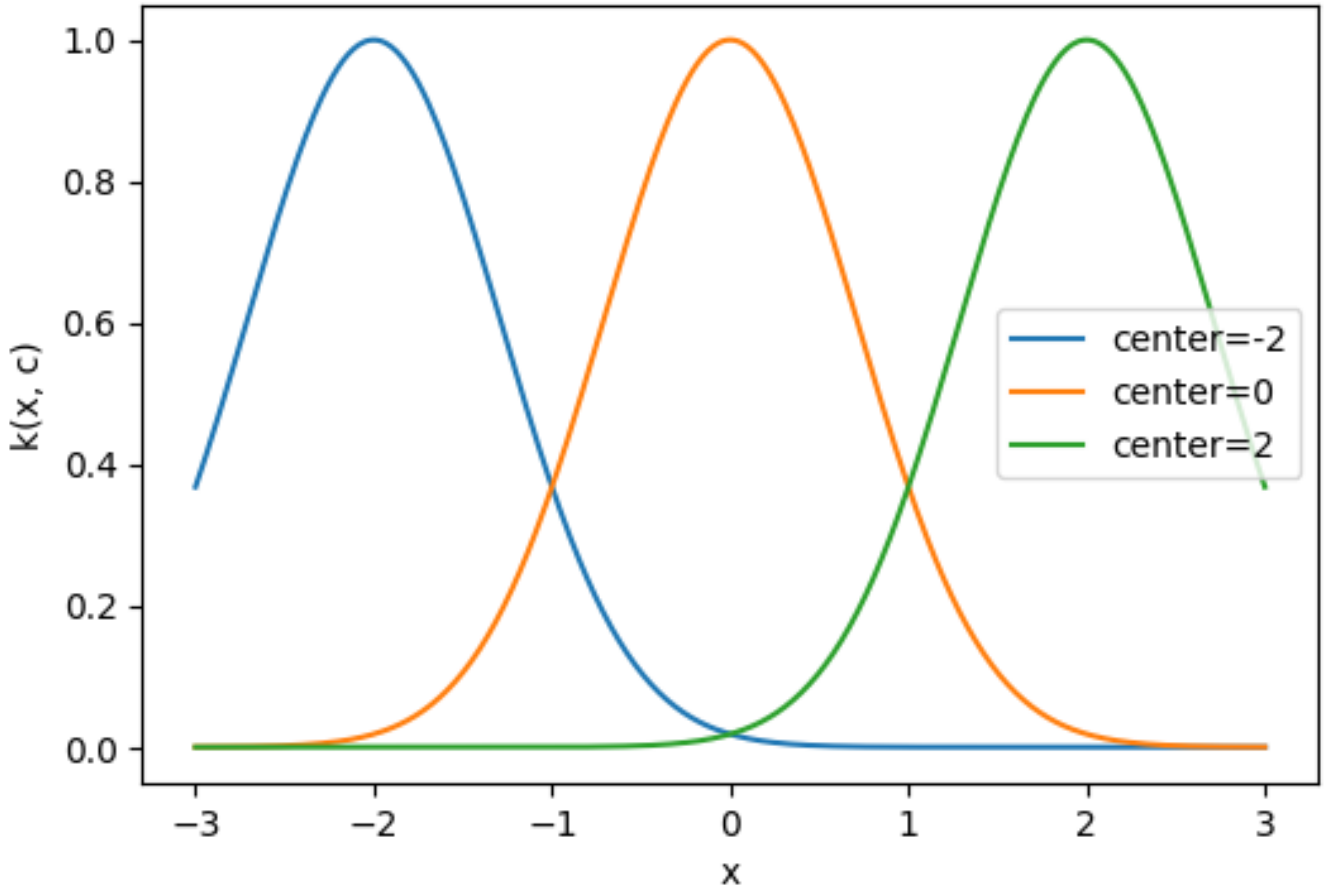
We can replace the function k with any suitable distance function. One of the most used is the Radial Basis Function:

$$k(x, x') = \exp(-\gamma \|x - x'\|^2)$$

Mental Health Mapping

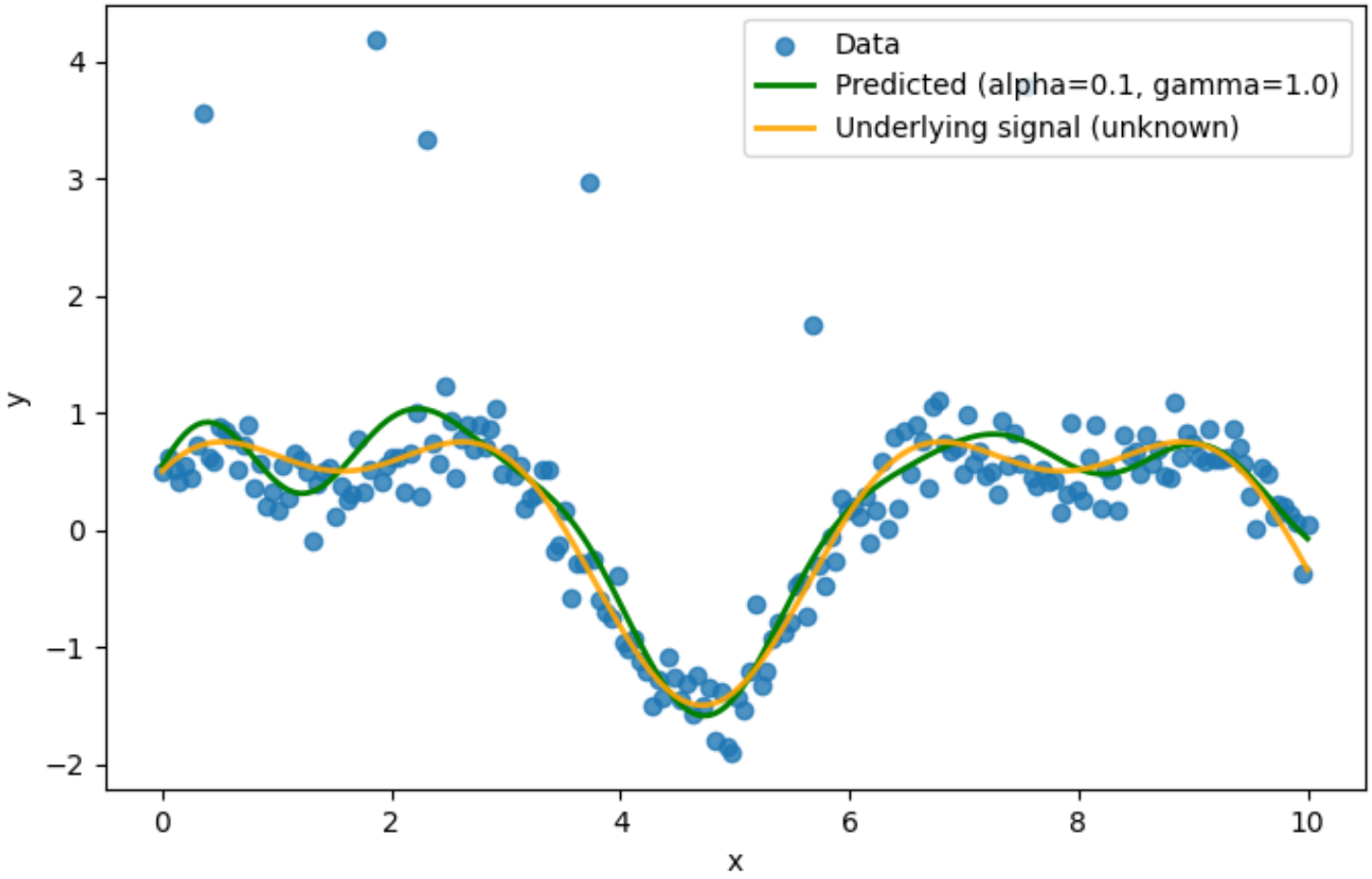
Regression: Kernel Ridge Regression

One can imagine the RBF kernel as taking a random point and introducing a gaussian centered (mean) in that point:



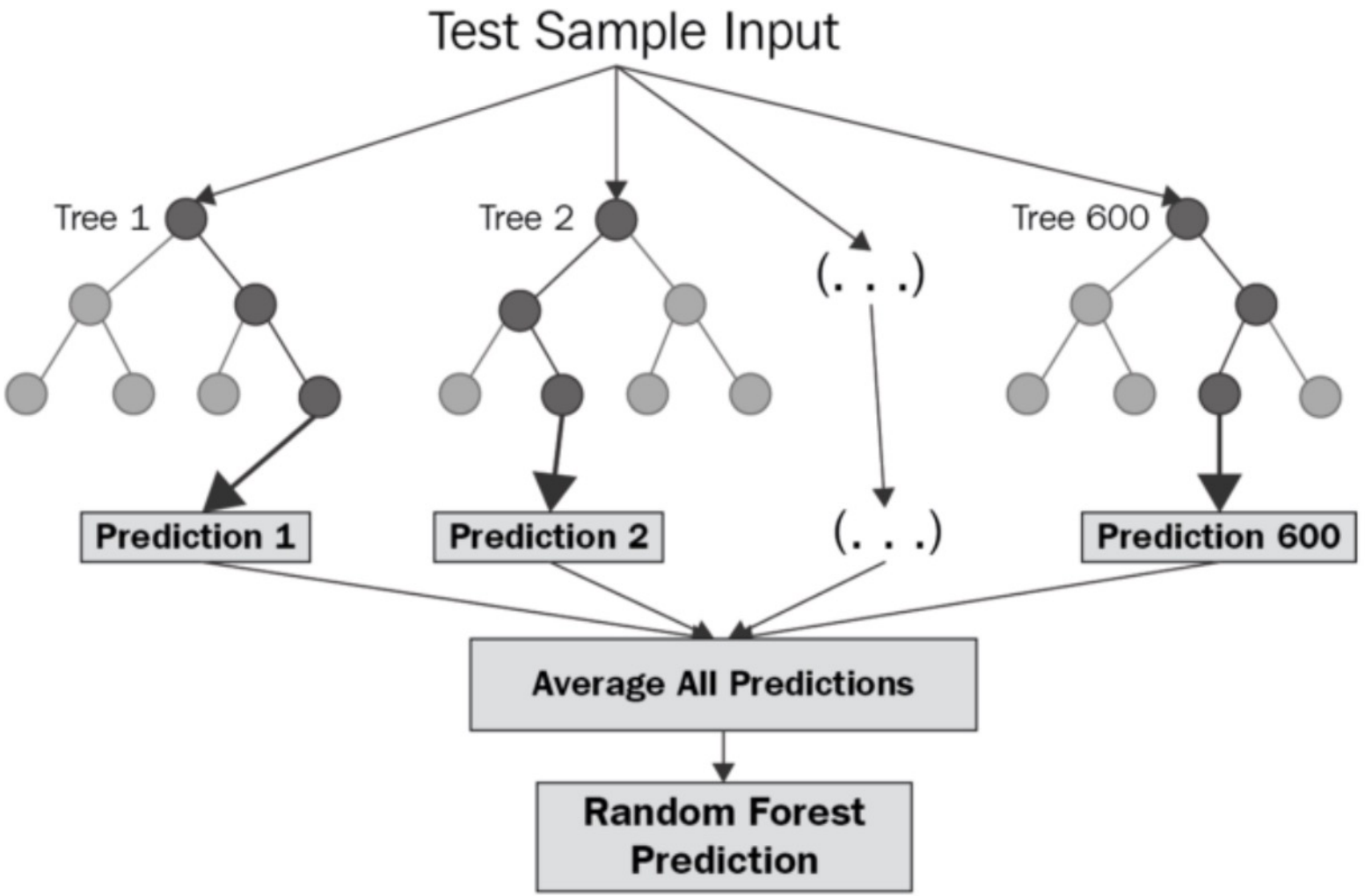
Regression: Kernel Ridge Regression

Kernel Ridge learns how to “weight” all these gaussian contribution for the training points and use this result to predict the next sample.



Mental Health Mapping

Regression: Random Forest



Mental Health Mapping

Explainability: XGBoost

$$\text{obj} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \omega(f_k)$$

$$\begin{aligned}
\hat{y}_i^{(0)} &= 0 \\
\hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\
\hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\
&\dots \\
\hat{y}_i^{(t)} &= \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)
\end{aligned}$$

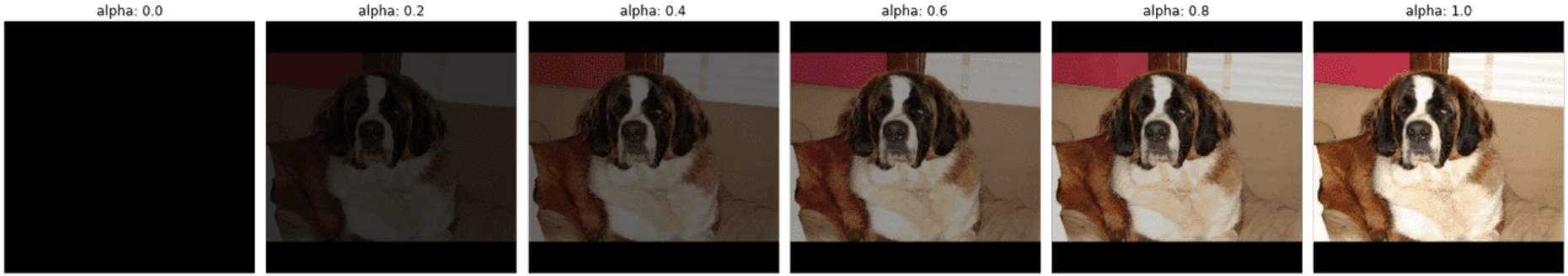
Here w is the vector of scores on leaves, q is a function assigning each data point to the corresponding leaf, and T is the number of leaves. In XGBoost, we define the complexity as

$$\omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

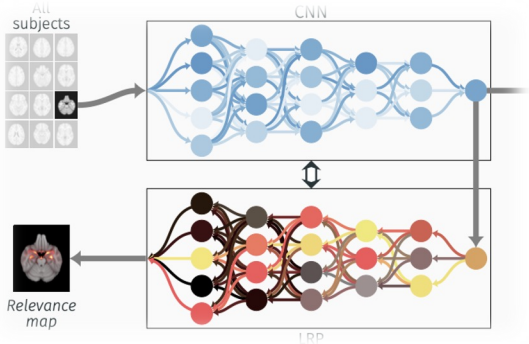
Mental Health Mapping

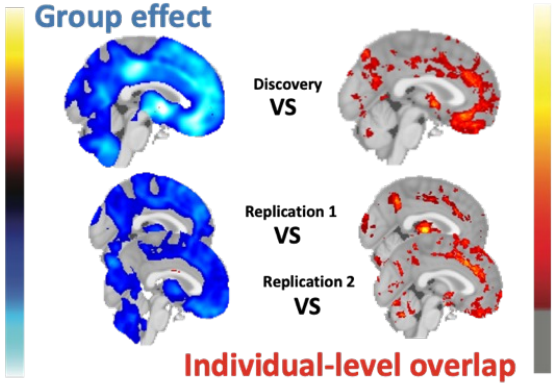
Explainability: Integrated Gradients

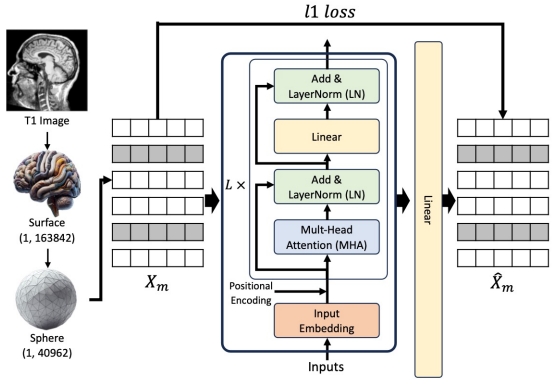
$$IntegratedGrads_i^{approx}(x) ::= (x_i - x_i') \times \sum_{k=1}^{m} \frac{\partial F(x' + \frac{k}{m} \times (x - x'))}{\partial x_i} \times \frac{1}{m}$$



Mental Health Mapping


All subjects
CNN
Relevance map
LRP


Group effect
Discovery VS
Replication 1 VS
Replication 2 VS
Individual-level overlap


l1 loss
T1 Image
Surface (1, 163842)
Sphere (1, 40962)
X_m
$L\times$
Add & LayerNorm (LN)
Linear
Add & LayerNorm (LN)
Mult-Head Attention (MHA)
Positional Encoding
Input Embedding
Inputs
Linear
$\hat{X}_m$

Thank you!


Mental Health Mapping